

# Feature Selection: Choosing What Matters

---

Nipun Batra and teaching staff

IIT Gandhinagar

July 31, 2025

# Outline

## 1. The Feature Selection Problem

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise



# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise
- **Computational cost:** Training and inference slow down

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise
- **Computational cost:** Training and inference slow down

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise
- **Computational cost:** Training and inference slow down
- **Storage cost:** More memory required

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise
- **Computational cost:** Training and inference slow down
- **Storage cost:** More memory required

# Why Feature Selection Matters

## The Challenge

**More features  $\neq$  Better performance**

**Problems with too many features:**

- **Curse of dimensionality:** Data becomes sparse
- **Overfitting:** Model memorizes noise
- **Computational cost:** Training and inference slow down
- **Storage cost:** More memory required

## Goal

Find the **minimal set** of features that gives **maximum predictive power**